Impact of Varying Panel Membership on Ratings of Appropriateness in Consensus Panels: A Comparison of a Multi- and Single Disciplinary Panel

Ian Coulter, Alan Adams, and Paul Shekelle

Objective. The objective of the study was to examine the appropriateness ratings for the use of spinal manipulation for low back pain of a multidisciplinary panel of medical and chiropractic physicians, and those of a panel composed only of chiropractic physicians.

Data Sources. The study analyzed data from two consensus panels conducted at RAND in 1990 and 1991.

Study Design. The study design followed that of the traditional RAND consensus panels. Nine individuals comprised each panel, and each panelist was asked to rate, on a nine-point scale, the indications for spinal manipulation twice, the first time alone and the second time jointly with the panel.

Data Collection. The ratings of the panelists from both groups, for both round one and round two, were collated and compared.

Principal Findings. While both panels were more likely to rate the indications as inappropriate than appropriate, the single disciplinary panel was more likely to rate an indication as appropriate than the multidisciplinary panel.

Conclusion. The composition of a panel clearly influences the ratings and those who use a given procedure in practice, in this case manipulation, are more likely to rate it as appropriate than those who do not use the procedure.

Key Words. Consensus panels, spinal manipulation, low back pain, chiropractic

An important issue that haunts the health care system is the necessity and appropriateness of the services provided to patients. Donabedian (1980) states:

> When care is said to be altogether unnecessary, or in some degree excessive, a judgment is made that the care is not expected to make, as a whole or in its parts, a contribution to the patient's health or welfare. (p. 6)

The key question for health professionals, however, is how to determine that a given service or course of treatment is unnecessary or appropriate. One

approach that has been used to address the issue of appropriate care is the consensus panel.

There is general agreement that the consensus panel method for resolving issues of the appropriateness of health care is an American invention (Vang 1986; Lomas 1991). Begun in 1977 by the National Institutes of Health, it was a method by which the scientific community could bring relevant research to bear on the quality of health care. In a sense, its purpose was to close the gap between what was known scientifically and what was practiced clinically. Since then, alternative approaches to the NIH model have arisen, both within the United States and in other countries.

Although variations of the consensus method have now been developed in other countries (Stocking 1985; Vang 1986; Casparie et al. 1987; Andreason 1988; Calltrop 1988), the two primary approaches in the United States are those developed at NIH and RAND. In both cases the ultimate target of the process is the clinician, and the common aim is to improve the quality of care. However, the way in which they go about these consensus-reaching processes is distinct.

Lomas (1991) in offering a framework for assessing the various types of consensus panels provides, at the same time, a taxonomy of sorts for the various approaches. The approaches may vary according to the choice of the topic, the choice of the members, the background preparation of the panelists, the way in which group judgments are derived, the criteria for determining recommendations, and the type of final report.

Of the various approaches, the RAND method has been the most extensively researched. Studies to date have investigated the relationship between the *literature* and the *ratings* (Fink et al. 1987); the reliability of the ratings (Park, Fink, Brook et al. 1986; Merrick, Fink, Park, et al. 1987; Brook, Kosecoff, Chassin, et al. 1988; Chassin 1989; Leape et al. 1992; Kahn et al. 1992); face and content validity (Merrick, Fink, Park, et al. 1987; Kahn, Park,

---

This research was a joint undertaking of RAND; Department of Veterans Affairs, West Los Angeles Section of General Internal Medicine, the Consortium for Chiropractic Research (CCR); and the Foundation for Chiropractic Education and Research (FCER).

Address correspondence and requests for prints to Ian Coulter, Ph.D., Health Consultant, RAND, 1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407–2138. Alan Adams, D.C. is Vice President for Professional Affairs, Los Angeles College of Chiropractic; and Paul Shekelle, M.D., Ph.D., is Assistant Professor of Medicine, UCLA, and Research Associate of the Department of Veterans Affairs, Health Services Research; and a Health Consultant at RAND. This article, submitted to *Health Services Research* on June 15, 1994, was revised and accepted for publication on February 6, 1995.

and Brook 1988; Chassin, Kosecoff, Park, et al. 1986; Chassin, Park, Fink, et al. 1986; ; Solomon, Brook, Fink, et al. 1986; Kahn, Roth, Kosekoff, et al. 1986; Hilborne, Leape, Kahan, et al. 1991; Shekelle, Adams, Chassin, et al. 1991a,b; Shekelle, Adams, Chassin, et al. 1992; Bernstein, Laouri, Hilborne, et al. 1992; Berstein, McGlynn, Kamberg, et al. 1992); and construct validity (Merrick, Fink, Park, et al. 1987; Chassin, Kosekoff, Park, et al 1989; Brook, Park, Chassin, et al. 1990; McClellan and Brook 1992). However, as Brook (1993) notes, much work is still needed to expand our knowledge about the validity and reliability of the process.

The objective of this article is to continue this work by analyzing the effect a variation in membership in the panel has on the outcome of the ratings of indications. In an earlier study Leape et al. (1992) contrasted the outcome of a seven-member, all-surgical panel with that of a nine-member, multidisciplinary panel (surgeons and nonsurgeons) for the appropriateness of carotid endarterectomy. As predicted, the surgeons (the "doers") found more indications appropriate and fewer inappropriate. Several interesting conclusions are possible from this study. First, those surgeons on the balanced panel had rating scores at an intermediate point between the all-surgeon panelists and the no-surgeon panelists; this would imply that the panel process modifies the ratings of the doers in a mixed panel. Second, single-specialty panels have higher agreement than multidisciplinary panels. There were several limitations in this study. The panel sizes were different (nine versus seven). The time lag between the panels was considerable (comparing care as delivered in 1981 versus 1988 care). While they note that no additional research trials occurred in the interim, it is difficult to know how much medical practice changed in that period or the effect of the dissemination of the results of the first panel on the second panel. Furthermore, the all-surgeon panel could not change indications and had a different definition of "agree" and "disagree."

THE STUDY

The RAND consensus panel method has been extensively described elsewhere (Brook, Chassin, Fink, et al. 1986; Park, Fink, Brook, et al. 1986; Chassin et al. 1986a,b; Kahn, Roth, Kosecoff, et al. 1986; Kahn, Park, and Brook 1988; Brook, Park, Chassin, et al. 1990). In 1990 and 1991 RAND convened two separate panels to consider the appropriateness of spinal manipulation for low back pain. Both of these panels have been described in the literature

(Shekelle, Adams, Chassin, et al. 1991a,b; 1992). The first panel, the multidisciplinary panel, was composed of nine members: two orthopedists, one osteopath, one internist, one family physician, three chiropractors, and one individual with both chiropractic and medical qualifications who practices as a neurologist. This panel had six members with academic appointments and three from private practice. Chiropractors comprised the second panel, five from chiropractic teaching institutions and four general practitioners.

The panelists were chosen because of their clinical expertise, community influence (in professional organizations, for example), and geographical location (to give a diversity of geographic representation from the four major census areas of the United States).

The initial set of indications was constructed by the research staff based on the literature review (Shekelle, Adams, Chassin, et al. 1991b, 1992) and discussions with health professionals practicing manipulation. The indications categorized patients in terms of their history, symptoms, physical and radiological findings, and response to prior treatment. The attempt was to create lists that were detailed, comprehensive, and manageable. Furthermore, the lists needed enough detail so that patients presenting with a particular indication would be relatively homogeneous, in that the procedure would be equally appropriate or inappropriate for all of them. The objective was to include all of the indications that might arise in a practice for doing the procedure for low back pain.

The indications were then organized into "chapters" that, in most cases, corresponded to major symptoms or primary problems. These were changed slightly in round two of the first panel. The second panel began with the indications of round two of the first panel. The chapter titles for the indications were:

1. Acute low back pain, no neurological findings
2. Acute low back pain, no neurological findings, but with sciatic nerve irritation
3. Acute low back pain, minor neurological findings, no sciatic nerve irritation
4. Acute low back pain, minor neurological findings, with sciatic nerve irritation
5. Acute low back pain
6. Subacute low back pain, no prior treatment
7. Subacute low back pain, prior manipulative treatment with favorable response

8. Chronic low back pain, no prior treatment
9. Chronic low back pain, prior nonmanipulative treatment with favorable response
10. Chronic low back pain, prior laminectomy
11. An otherwise appropriate indication and:
    A. Possible clotting disorder
    B. Possible abdominal aortic aneurysm
    C. Definite abdominal aortic aneurysm, by history and imaging
    D. Repaired abdominal aortic aneurysm
    E. Spondylolisthesis

This approach generated 1,577 indications for spinal manipulation. Both panels received the same literature review (Shekelle, Adams, Chassin, et al. 1991a) and both were asked to rate the indications in the same way on an appropriateness scale of 1 to 9.

Each panel was provided with a set of definitions to accompany the ratings. There were two opportunities for these definitions (and the format of the ratings) to be altered. Following the circulation of the definitions/indications to the panelists individually, the panelists during the panel meetings could recommend changes. Very few changes were made. In the area of definitions, the changes were minor and had to do with clarification. The all-chiropractic panel made a minor modification to the definition of chiropractic manipulation. According to the multidisciplinary panel, "chiropractic treatment uses spinal manipulation along with other measures, such as mobilization, traction techniques, and massage." This was altered by the chiropractic panel to read "along with other measures, such as flexion-distraction, mobilization. . . ." The definition of HNP was altered from "HNP means herniated nucleus pulposus" to "HNP means herniated (not bulging) nucleus pulposus." Joint dysfunction, defined in the multidisciplinary panel as "decreased or aberrant joint mobility excluding hypermobility" and including the footnote "Not all panelists agreed that this is a clinically definable entity," was redefined as "decreased or aberrant segmental or sectional joint mobility excluding hypermobility but including hypertonic contraction of the paraspinous muscles" by the all-chiropractic panel. These changes in definitions are minor and not sufficient to greatly affect the outcome of the ratings. The all-chiropractic panel did alter the structure of some of the indications. In addition to some minor changes affecting a small number of indications in four chapters, they also deleted the difference between their interpretation of time since the onset of pain and that of the multidisciplinary panel in Chapter 8, and split

biomechanical and psychological stress in Chapters 9 and 10. Overall, 1,211 indications (78 percent) were unchanged between the two panels. The final round of ratings of both panels were analyzed using identical definitions of "agreement," "disagreement," "appropriate," and "inappropriate" as previously described (Shekelle, Adams, Chassin, et al. 1991a, 1992).

With regard to the indications, this article focuses only on those indications for which no changes were made between the two panels (1,211 indications).

AIMS

Based on earlier research some hypotheses can be made with respect to the ratings of the two panels.

Hypothesis 1. *Those who utilize a procedure in their practice are more likely to rate a higher proportion of indications for the procedure as appropriate, and a lower proportion inappropriate, than those who do not use the procedure.*

Hypothesis 2. *Those who utilize a procedure in their practice are more likely to have agreement about the appropriateness or inappropriateness of the procedure and to have less uncertainty than those who do not.*

RESULTS

HYPOTHESIS 1

The total of 1,211 agreed-upon indications represented 78 percent of the initial indications. The distribution of the ratings for the two panels is given in Table 1. This indicates that the chiropractors were more likely to rate an indication as appropriate[1] than was the multidisciplinary panel (33 percent of the indications versus 9 percent), while the latter rated more indications as inappropriate compared to the chiropractors (54 percent versus 45 percent). Table 2 indicates that the two panels agreed on 44 percent of the indications (those in the diagonal of the table). On 14 percent of the indications (those above the diagonal) the multidisciplinary panel was less conservative than the all-chiropractic panel, and in 42 percent (those below the diagonal) the multidisciplinary panel is more conservative. By conservative we mean that the panel is more likely to rate the indication as uncertain or inappropriate. The all-chiropractic panel was more likely to rate an indication as appropriate

and less likely to rate it as inappropriate than the multidisciplinary panel. However, it should be noted that even the chiropractic panel was more likely to rate an indication inappropriate than appropriate (45 percent versus 33 percent). These results support earlier research in showing more agreement between the two panels with regard to inappropriate indications than appropriate indications.

To test whether this is a general tendency by the all-chiropractic panel, we can examine the results by chapter (see Table 3). Table 3 shows that while the level of agreement does vary considerably by chapter (from 22 percent to 78 percent) with the exception of Chapter 1, the all-chiropractic panel always chooses more indications as appropriate than the multidisciplinary panel. But for inappropriate indications, the multidisciplinary panel considered more indications to be inappropriate than the chiropractic group–with the exception of Chapters 1, 3, and 7. Chapter 1 is the only chapter in which the multidisciplinary panel rated more indications as appropriate than inappropriate. The chiropractors for their part did this on only three of the chapters. The tendency for both groups, therefore, is to rate more indications as inappropriate than appropriate.

The disagreement between the two panels can be in two directions: first, where the multidisciplinary panel sees the indication as appropriate and the chiropractors see it as uncertain or inappropriate; second, where the chiropractors see the indication as appropriate and the multidisciplinary group

Table 1: Number and Percentage of Response of the Two Panels

	Response					
	Appropriate		*Uncertain*		*Inappropriate*	
Multidisciplinary	109	(9%)	447	(36.9%)	655	(54.1%)
All chiropractic	396	(32.7%)	265	(21.8%)	550	(45.4%)

Table 2: Comparison of the Multidisciplinary and All-Chiropractic Panel over All the Ratings

		All-Chiropractic Panel							
		Appropriate A		*Uncertain U*		*Inappropriate I*		*Total*	
Multidisciplinary panel	A	80	(6.6%)	3	(.25%)	26	(2.1%)	109	(9.0%)
	U	229	(19%)	71	(5.9%)	147	(12%)	447	(37%)
	I	87	(7.2%)	191	(15.8%)	377	(31.1%)	655	(54%)
Total		396	(32.7%)	265	(21.9%)	550	(45.4%)	1211	

Table 3: Comparison of the Two Panels by Chapter

	Response				
	Appropriate		*Inappropriate*		
Chapter	*Multi*	*Chiro*	*Multi*	*Chiro*	*Agreement*
1	69	63	20	51	54%
2	0	60	64	51	22%
3	37	53	40	48	41%
4	0	45	89	48	25%
5	4	58	125	108	46%
6	3	69	228	162	56%
7	1	21	33	36	41%
8	1	12	22	20	50%
9	0	8	26	18	58%
10	0	1	7	6	78%

Table 4: Comparison of the Two Panels by Disagreement

		Chiropractic Panel		
		Appropriate	*Uncertain*	*Inappropriate*
Multidisciplinary panel	A	–	3	26
	U	229	–	147
	I	87	191	–

sees it as either uncertain or inappropriate. These data are shown in Table 4. The largest disagreement occurs when the chiropractors rate an indication as appropriate and the multidisciplinary panel rate it as uncertain, followed by the situation where the chiropractors are uncertain and the multidisciplinary panel rates the indication as inappropriate. While this is a result predicted by Hypothesis 1, Table 4 also shows that extreme disagreements are much less common. On only 26 indications (3.8 percent) does the multidisciplinary panel rate something as appropriate that the all-chiropractic panel rates as inappropriate, and in only 87 instances (12.7 percent) was the reverse true. It would appear that the "gray areas" lead to the most disagreement.

Where the chiropractors were likely to rate an indication as appropriate and the multidisciplinary panel as inappropriate, the clinical characteristics of the indications involved sciatic nerve irritation, pain no longer present, central herniated nucleus pulposus, spinal stenosis or free fragment, and prior response or unfavorable response. Where the multidisciplinary panel rated an indication as appropriate but the chiropractors rated it as inappropriate,

all but one of the disagreements occurred in acute pain chapters and were indications without joint dysfunctions. (Similarly, when the multidisciplinary panel chose "appropriate" and the chiropractors chose "uncertain," it was around the clinical issue of joint dysfunction.) Hypothesis 1 may also be further tested by examining the ratings of the chiropractors on the multidisciplinary panel. The hypotheses would predict that they would rate more indications as appropriate and that their ratings should be similar to those of the all-chiropractic panel. If these ratings are dissimilar and more closely allied to those of the panel members who do not perform manipulation, this suggests that the panel process itself leads to a modification of the ratings of those who perform a procedure. Table 5 presents the comparison of the ratings of the chiropractors on the mixed panel to both the chiropractic panel and the overall ratings of the mixed panel. As the data in Table 5 show, the chiropractors on the mixed panel rated 20.4 percent of the indications as appropriate compared to 27 percent by the all-chiropractic panel and only 7 percent by the multidisciplinary (mixed) panel. Overall, therefore, our first hypothesis is supported by the data.

HYPOTHESIS 2

The second hypothesis can be tested by looking at the levels of consensus within the two panels. Table 6 gives the data for the multidisciplinary panel; the data show that in both the initial round and the final round the all-chiropractic panel had the lowest percentage of disagreement and the highest percentage of agreement. Furthermore, their mean absolute deviation from the median was lower in both rounds. Further evidence is seen in the percentage of indications found inappropriate, uncertain, and appropriate for each panel (see Table 7). Despite the differences between the two panels however, the over-all direction of their ratings is similar. That is, for most of the chapters the tendency of both groups is to rate more indications as inappropriate than

Table 5: Comparison of Appropriateness Ratings of the All-Chiropractic Panel and the Chiropractors on the Multidisciplinary Panel

	Multidisciplinary Panel	*Chiropractors on Multidisciplinary Panel*	*Chiropractic Panel*
Number of appropriate ratings	112	317	425
Percentage of appropriate ratings	7%	21%	27%

appropriate (the multidisciplinary group did this for all but one chapter and the all chiropractic panel on all but three chapters).

The single disciplinary panel therefore not only had a higher degree of agreement, they also had a higher degree of agreement with regard to doing spinal manipulation as measured by the greater number of indications they rated as appropriate. With regard to the uncertain ratings, the difference between the two groups is not so great. Furthermore, they are closer with regard to inappropriate ratings. The data therefore support the second hypothesis.

DISCUSSION

The results presented here are in line with earlier studies on the panel process. Leape et al. (1992) have already shown that single disciplinary panels are more likely to rate an indication as appropriate than are multidisciplinary panels and, conversely, to rate fewer indications as inappropriate. However, both studies confirm that these panels are more likely to rate something as inappropriate than appropriate. These findings are similar to those of Brook, Kosecoff,

Table 6: Median Ratings and Extent of Agreement and Disagreement on Appropriateness Ratings for the Multidisciplinary and the All-Chiropractic Panels

	Initial Ratings		*Final Ratings*	
Item	*Multi*	*Chiro*	*Multi*	*Chiro*
Number of indications	1577	1550	1550	1570
Average median	3.34	4.61	3.16	3.95
Mean absolute deviation from the mean	1.70	1.39	1.14	0.83
Percentage of agreement	11.79	27.23	35.74	63.18
Percentage of disagreement	36.84	11.55	11.90	8.10

Table 7: Categories of Appropriateness for Indications for Spinal Manipulation for the Multidisciplinary and All-Chiropractic Panels

	Number of Indications		*Percentage of Indications*	
Category	*Multi*	*Chiro*	*Multi*	*Chiro*
Inappropriate	924*	750	60	48
Uncertain	514	395	33	25
Appropriate	112	425	7	27

*The two panels rated a different number of indications.

Chassin, et al. (1988) in a comparison of two panels rating indications for coronary disease, one in the United States and one in the United Kingdom. Here, although the percentages for appropriate ratings differed, the two panels tended to rate the appropriateness of the indications in the same order. More recently, Frazer et al. (1994) have shown with cholecystectomy that an all-surgical panel agrees on more indications and fewer contraindications than a mixed panel. Differences have also been found in panels in Britain and Israel.

As noted earlier in this article, the significant disagreements between the two panels, where one panel rated an indication as appropriate and the other as inappropriate (and conversely), all occurred, except one, in acute pain chapters and with regard to indications without joint dysfunction. This reflects the fact that joint dysfunction had a very different significance for the chiropractic panel than it did for the multidisciplinary panel (many of whom thought joint dysfunction is not a clinical entity). For chiropractors joint dysfunction operates as virtually a necessary condition for rating any indication as appropriate (the *sine qua non* of the manipulable lesion). With the exception of those who manipulate on the multidisciplinary panel, the non-chiropractic panelists were less convinced of its importance. Consequently, their ratings were more stable around that variable. The chiropractors on both panels would move their ratings dramatically based on the variable of joint dysfunction. These results suggest that a fruitful analysis for the future will be to examine each panel for "critical factors." Such an analysis would provide a picture of what is felt to be clinically important in either panel and, in this regard, point out the contrast between the chiropractors and medical doctors.

The other major conclusion from this article is that the composition of the panel clearly influences the ratings. Those who do not use a procedure, such as manipulation, are more conservative in what they will rate as appropriate for the procedure, and are less likely to be in agreement about this issue than those who use the procedure in their practice. Further, an increase in the percentage of indications rated appropriate by a panel may have a drastic effect on the measurement of appropriate and inappropriate care. In the study by Leape et al. (1992), an additional 10 percent of indications rated appropriate resulted in an increase from 38 percent, based on the mixed panel, to 70 percent based on the specialist panel, in the ratings of the patient files.

Depending on how one views the purpose of a panel, and whether the panel should err on the side of caution, this is either an argument in favor of multidisciplinary panels or an argument against them. RAND traditionally has favored multidisciplinary panels. The logic for this has been that the

spectrum of disease is unlikely to be experienced by a single category of practitioner. Further, the stage at which they encounter the disease will vary considerably from general to specialty practice. Certain types of patients will present to different types of physicians. Furthermore, those who perform a procedure have both a professional bias in favor of it (if not, presumably they would not be performing it) and a financial incentive to support its use. The results presented here contribute additional information on the effect of a variation in panel membership for two panels whose recommended indications were more similar than in previous studies. These results will contribute further to the debate over the composition of panels.

In the absence of a gold standard, however, it is not possible to judge whether a specialist or a mixed panel is better. The most we can say is that they differ. Two areas of further research may partially resolve this problem. In the first, the ratings can be compared to the literature. For spinal manipulation there is now a sufficient number of clinical studies (including random controlled trials) to allow a meta-analysis (Shekelle, Adams, Chassin, et al. 1992). A second possibility is to compare the ratings to the newly published AHCPR guidelines for treatment of low back pain (Bigos, Bowyer, and Braen 1994).

Ultimately, however, the important comparison will be between panel ratings of appropriateness and necessity and patient outcomes. Until this is possible, mixed panels will be preferable when the purpose is coming to consensus about overuse of a procedure. The result here will be a reduction in inappropriate care. However, in those instances where the problem is underuse of a procedure, as measured by necessity, a mixed panel could have the opposite effect and further contribute to underuse.

REFERENCES

Andreasen, P. B. 1988. "Consensus Conferences in Different Countries." *International Journal of Technology Assessment in Health Care* 4 (2): 305–308.

Bernstein, S. J., M. Laouri, L. H. Hilborne, L. L. Leape, J. P. Kahan, R. E. Park, C. J. Kamberg, and R. H. Brook. 1992. *Coronary Angiography: A Literature Review and Ratings of Appropriateness and Necessity.* Publication No. JRA-03-CWF/HF. Santa Monica, CA: RAND.

Bernstein, S. J., E. A. McGlynn, C. J. Kamberg, M. R. Chassin, G. A. Goldberg, A. L. Siu, and R. H. Brook. 1992. *Hysterectomy: A Literature Review and Ratings of Appropriateness.* Publication No. JR-04. Santa Monica, CA: RAND.

Bigos, S., D. Bowyer, and S. Braen. 1994. "Acute Low Back Problems in Adults: Clinical Practice Guidelines." *Quick Reference Guide Number 14.* U.S. Public Health Service, Publication No. (AHCPR) 95–0643. Rockville, MD: Department of Health and Human Services, December.

Brook, R. H., M. R. Chassin, A. Fink, D. H. Solomon, J. Kosecoff, and R. E. Park. 1986. "A Method for the Detailed Assessment of the Appropriateness of Medical Technologies." *International Journal of Technology Assessment in Health Care* 2 (1): 53–63.

Brook, R. H., J. B. Kosecoff, M. R. Chassin, R. E. Park, C. M. Winslow, and J. R. Hampton. 1988. "Diagnosis and Treatment of Coronary Disease: Comparison of Doctors' Attitudes in the USA and the UK." *Lancet* 1 (April): 750–53.

Brook, R. H., R. E. Park, M. R. Chassin, D. H. Solomon, J. Keesey, and J. Kosecoff. 1990. "Predicting the Appropriate Use of Carotid Endarterectomy, Upper Gastrointestinal Endoscopy, and Coronary Angiography." *The New England Journal of Medicine* 323 (17): 1173–77.

Brook, R. H. 1993. "The RAND/UCLA 'Appropriateness Method'." In *Clinical Practice Guideline Development: Methodologic Perspectives*. AHCPR, Rockville, MD: Agency for Health Care Policy and Research, Public Health Service.

Calltrop, J. 1988. "Consensus Development Conferences in Sweden." *International Journal of Technology Assessment in Health Care* 4 (1): 75–88.

Casparie, A. F., N. S. Klazinga, J. J. E. van Everdingen, and P. J. Touw. 1987. "Health-Care Providers Resolve Clinical Controversies: The Dutch Consensus Approach." *Australian Clinical Review* (March): 43–47.

Chassin, M. R., J. Kosecoff, R. E. Park, A. Fink, S. Rauchman, J. W. Keesey, and R. H. Brook. 1986. *Indications for Selected Medical and Surgical Procedures–A Literature Review and Ratings of Appropriateness: Coronary Angiography.* Publication No. R-3204/1-CWF/HF/HCFA/PMT/RWJ. Santa Monica, CA: RAND.

Chassin, M. R., J. Kosecoff, R. E. Park, C. M. Winslow, K. Kahn, N. Merrick, A. Fink, J. Keesey, R. H. Solomon, and R. H. Brook. 1989. *The Appropriateness of Use of Selected Medical and Surgical Procedures and Its Relationship to Geographic Variations in Their Use.* Ann Arbor, MI: Health Administration Press.

Chassin, M. R., R. E. Park, A. Fink, S. Rauchman, J. W. Keesey, and R. H. Brook. 1986. *Indications for Selected Medical and Surgical Procedures–A Literature Review and Ratings of Appropriateness: Coronary Artery Bypass Graft Surgery.* Publication No. R-3204/2-CWF/HF/HCFA/PMT/RWJ. Santa Monica, CA: RAND.

Chassin, M. 1989. "How Do We Decide Whether an Investigation or Procedure Is Appropriate?" In *Appropriate Investigation and Treatment in Clinical Practice*, ed. by A. Hopkins. London: Royal College of Physicians.

Donabedian, A. 1980. *Explorations in Quality Assessment and Monitoring*, Vol. I, II, III. Ann Arbor, MI: Health Administration Press.

Fink, A., R. H. Brook, J. Kosecoff, and M. R. Chassin. 1987. "Sufficiency of the Clinical Literature for Learning about the Appropriate Uses of Six Medical and Surgical Procedures." *Western Journal of Medicine* 147 (5): 609–15.

Frazer, S. M., D. Pilpel, J. Kosecoff, and R. H. Brook. 1994. "Effect of Panel Composition on Appropriateness Ratings." *International Journal for Quality in Health Care* 6 (3): 251–55.

Hilborne, L. H., L. L. Leape, J. P. Kahan, R. E. Park, C. J. Kamberg, and R. H. Brook. 1991. *Percutaneous Transluminal Coronary Angioplasty: A Literature Review and Ratings of Appropriateness and Necessity.* Publication No. JRA-01-CWF/HF. Santa Monica, CA: RAND.

Kahn, K. L., R. E. Park, and R. H. Brook. 1988. "The Effect of Comorbidity on Appropriateness Ratings for Two Gastrointestinal Procedures." *Clinical Epidemiology* 41 (2): 115–22.

Kahn, K. L., R. E. Park, J. Vennes, and R. H. Brook. 1992. "Assigning Appropriateness Ratings for Diagnostic Upper Gastrointestinal Endoscopy Using Two Different Approaches." *Medical Care* 30 (11): 1016–28.

Kahn, K. L., C. P. Roth, J. Kosecoff, A. Fink, R. H. Brook, M. R. Chassin, D. H. Solomon, R. E. Park, J. W. Keesey, and M. F. Flynn. 1986. *Indications for Selected Medical and Surgical Procedures–A Literature Review and Ratings of Appropriateness: Diagnostic Upper Gastrointestinal Endoscopy.* Publication No. R-3204/4-CWF/HF/HCFA/PMT/RWJ. Santa Monica, CA: RAND.

Leape, L. L., R. E. Park, J. P. Kahan, and R. H. Brook. 1992. "Group Judgments of Appropriateness: The Effect of Panel Composition." *Quality Assurance in Health Care* 4 (2): 151–59.

Lomas, J. 1991. "Words Without Action? The Production, Dissemination and Impact of Consensus Recommendations." *Annual Review of Public Health* 12: 42–65.

McClellan, M., and R. H. Brook. 1992. "Appropriateness of Care Comparison of Global and Outcome Methods to Set Standards." *Medical Care* 30 (7): 565–86.

Merrick, N. J., R. H. Brook, R. E. Park, J. Kosecoff, C. F. Roth, L. McCloskey, J. W. Keesey, D. H. Solomon, and A. Fink. 1986. *Indications for Selected Medical and Surgical Procedures–A Literature Review and Ratings of Appropriateness: Carotid Endarterectomy.* Publication No. R-3204/6-CWF/HF/HCFA/PMT/RWJ. Santa Monica, CA: RAND.

Merrick, N. J., A. Fink, R. E. Park, R. H. Brook, J. Kosecoff, M. R. Chassin, and D. H. Solomon. 1987. "Derivation of Clinical Indications for Carotid Endarterectomy by an Expert Panel." *American Journal of Public Health* 77 (2): 187–90.

Park, R. E., A. Fink, R. H. Brook, M. R. Chassin, R. L. Kahn, N. J. Merrick, J. Kosecoff, and D. H. Solomon. 1986. "Physician Ratings of Appropriate Indications for Six Medical and Surgical Procedures." *American Journal of Public Health* 76 (7): 766–72.

Shekelle, P. G., A. H. Adams, M. R. Chassin, E. L. Hurwitz, and R. H. Brook. 1992. "Spinal Manipulation for Low-Back Pain." *Annals of Internal Medicine* 117 (7): 590–98.

Shekelle, P. G., A. H. Adams, M. R. Chassin, E. L. Hurwitz, R. E. Park, R. B. Phillips, and R. H. Brook. 1991a. *The Appropriateness of Spinal Manipulation of Low-Back Pain: Indications and Ratings by a Multidisciplinary Expert Panel.* Publication No. R-4025/2-CCR/FCER. Santa Monica, CA: RAND.

Shekelle, P. G., A. H. Adams, M. R. Chassin, F. L. Hurwitz, R. B. Phillips, and R. H. Brook. 1991b. *The Appropriateness of Spinal Manipulation of Low-Back Pain: Project Overview and Literature Review.* Publication No. R-4025/1-CCR/FCER. Santa Monica, CA: RAND.

——. 1992. *The Appropriateness of Spinal Manipulation of Low-Back Pain: Indications and Ratings by an All Chiropractic Expert Panel.* Publication No. R-4025/3-CCR/FCER. Santa Monica: RAND.

Solomon, D. H., R. H. Brook, A. Fink, R. E. Park, and J. W. Keesey. 1986. *Indications for Selected Medical and Surgical Procedures–A Literature Review and Ratings of Appropriateness: Cholecystectomy.* Publication No. R-3204/3-CWF/HF/HCFA/PMT/RWJ. Santa Monica, CA: RAND.

Stocking, B. 1985. "First Consensus Development Conference in the United Kingdom: On Coronary Artery Bypass Grafting." *British Medical Journal* 291 (14 September): 713–18.

Vang, J. 1986. "The Consensus Development Conference and the European Experience." *International Journal of Technology Assessment in Health Care* 2 (1): 65–76.